# 基于字词向量的多尺度卷积神经网络微博评论的情感分类模型及实验研究*

■ 张柳[1] 王晰巍[1,2] 黄博[3] 刘宇桐[1]
[1] 吉林大学管理学院 长春 130022 [2] 吉林大学大数据管理研究中心 长春 130022
[3] 吉林大学计算机科学与技术学院 长春 130022

**摘要：**[目的/意义]微博评论情感分类模型可以为相关舆情监管部门正确管控话题事件的发展状况和舆情提供一定的指导作用。[方法/过程]基于字词向量的多尺度卷积神经网络，运用多尺度卷积核改善微博评论中上下文信息有限的条件制约，构建基于字词向量的多尺度卷积神经网络微博评论情感分类模型；通过爬取“微博热搜整改”数据，对模型的可行性和优越性进行验证。[结果/结论]验证结果表明基于字词向量的多尺度卷积神经网络在微博舆情等上下文信息有限的短文本分类任务中表现良好。本文在理论层面为微博舆情情感分类提供了更为准确的情感分类理论模型及分类方法，在实践层面可以更好地指导舆情监管部门对舆情的情感倾向进行更好的引导和监管。

**关键词：**卷积神经网络 字词向量 多尺度卷积核 微博评论 情感分类
**分类号：**G203
**DOI**：10.13266/j.issn.0252-3116.2019.18.012



随着社交媒体的快速发展，越来越多的网络用户通过社交媒体平台表达自己的情感、观点和看法。根据2018年第二季度微博财报数据显示，截至2018年6月微博月活跃用户共4.31亿，与上一季度相比增长了2 100万，移动端占比达到93%，日活跃用户数增至1.9亿[1]。微博作为一种新兴的信息发布及社交媒体平台，具有开放性、即时性、交互性和自由性等特点，移动终端APP提供的便利性使得用户可以随时表达自己观点和情感，对微博评论进行情感分类成为学术界关注的新课题，它可以更好地帮助舆情管理部门进行舆情的引导和管理。

近年来国内外学者相继展开了舆情情感分类的相关研究。国外学者Rui N S[2]构建了聚合情感语义识别和观点传播影响模型，并着重研究了全球情感计算框架，分析了影响情感传播的影响因素；Soleymani M[3]定义了多模态情感分析问题，认为多模态情感分析具有重要开发潜力；Özturk N[4]利用Twitter收集23 881个公众对叙利亚难民危机看法的相关推文，并进行情感分析比较，数据结果表明土耳其语推文的情感与英文推文的情感具有很大的差别。国内学者唐晓波[5]构建特征本体对特征词进行分类，计算微博产品评论的特征情感极性和强度，并通过采集微博数据验证方法的有效性；马松岳[6]通过抓取豆瓣电影的用户评论，使用ROST EA工具进行情感分析，并通过回归方法构建模型进行检验；梁晓敏[7]构建了舆论对象分析模型并对舆论对象的关系网络进行研究，实验结果表明模型能有效识别主要舆论对象及其情感词，并能直观地展现网民对舆论对象随时间演化的情感表达和关系认知。

目前，情感分类这一类自然语言处理任务，一般被当做时序信号进行处理。其性能主要依赖于文本特征的表示方法。唐慧丰[8]提出采用BiGrams的特征表示方法，结合信息增益特征选择方法和支持向量机分类器，在训练集充足、特征数量适当的情况下表现良好。张冬雯[9]结合Word2Vec和SVMperf两类工具，采用扩充情感词典的方法，取得了很好的效果。

* 本文系国家自然科学面上项目“信息生态视角下新媒体信息消费行为机理及服务模式创新研究”(项目编号:71673108)研究成果之一。
作者简介：张柳(ORCID:0000-0001-8688-4959)，博士研究生，E-mail:598837913@qq.com；王晰巍(ORCID:0000-0002-5850-0126)，副院长，教授，博士生导师；黄博(ORCID:0000-0001-9128-4659)，博士研究生；刘宇彤(ORCID:0000-0003-3320-7369)，硕士研究生。
收稿日期:2018-12-15 修回日期:2019-02-23 本文起止页码:99-108 本文责任编辑:杜杏叶

图书情报工作
第 63 卷 第 18 期 2019 年 9 月

然而,基于文本特征的标识方法,多依赖于上下文的时序信息。由于微博评论的上下文信息十分有限,并含有大量的集外词(Out of Vocabulary,OOV),因而微博评论词级别向量很大程度上受中文分词的影响,无法提供足够的先验知识进行准确的情感分类,导致对网络微博舆情的情感分类不准确[10]。

近年来,不断有学者提出用图像领域的卷积神经网络来解决自然语言处理任务。冯兴杰[11]提出将卷积神经网络与注意力模型相结合,避免了人工构建情感词典的依赖,从而提升了模型整体的自动化性能。被誉为“卷积神经网络之父”的 Yann Lecun[12]提出,用 70 种字符粒度的向量(26 个英文字母,0 – 9 阿拉伯数字及常用符号),结合卷积神经网络,来进行文本分类。依赖于卷积层的特征提取能力,将这些字符粒度的向量提取为有意义的特征,取得了广泛关注。使得卷积神经网络不但在图像领域大放异彩,也提供了自然语言处理任务的新思路。

受此启发,本文基于卷积神经网络模型,选用两种粒度的向量——字级别以及词级别。将这两种向量拼接为三维向量矩阵,以此构建基于字词向量的多尺度卷积神经网络模型。本文在研究中试图解决以下三个方面的研究问题:①基于字词向量的多尺度卷积神经网络构建微博评论情感分类模型;②通过典型话题对构建的微博评论情感分类准确性进行验证分析;③验证所构建的微博评论情感分类模型的优越性。



## 1 相关理论

### 1.1 舆情情感分类

情感是人类对客观事件或对象所持有的态度和心理感受,它反映的是具有一定需求或愿望、观点的主体与客体之间的关系[13]。微博评论的情感分类,是指基于微博评论进行挖掘和分析,识别出其中的情感倾向性,并将情感文本分类为褒义、贬义评价,或依照情感强度进行分类、打分等相关任务[14]。由于微博评论的情感分类文本存在一定的时序性,通过引入注意力机制[15],可以将文本中不同部分对分类任务的贡献赋予不同的权重,从而训练权重矩阵。通过将循环神经网络中的 LSTM 模型与注意力机制相结合[16],可以选取微博评论作为数据源,针对热点问题的用户情感进行研究。

目前主流的舆情情感分类的方法[17],分为以下四类:

(1)基于词典的情感分类方法。此类方法的核心是“词典 + 规则”,即以情感词典为判断情感极性的主要依据,同时考量评论中的句法结构。此类方法的性能取决于情感词典以及相应判断规则的质量,并且此二者都需要人工干预,导致此类方法依赖于先验知识,泛化能力较差。

(2)基于机器学习的情感分类方法。此类方法的核心工作为特征工程,常用到的有 N-gram 特征,句法特征,词频逆文档频率(TF-IDF)特征等。此类方法使用经典的分类模型:如朴素贝叶斯,支持向量机等。同样,此类方法依然依赖于人工的特征设计,泛化能力有限。

(3)词典与机器学习混合的分类方法。此类方法将“词典 + 规则”视为一种特殊的特征,与现有特征(N-gram,句法特征等)进行结合,然后选取最终特征组合,再用分类器进行分类。此种方法结合而两者的优点,是目前小规模样本上最常用的方法。

(4)基于深度学习的方法。此类方法先从海量评论语料中训练出语义词向量,在通过不同的语义合成方法用训练出的词向量构建整体评论的特征表达。此类方法在数据充足的情况下表现良好,但在数据量有限的情况下无法学习足够的上下文信息。

### 1.2 中文分词与词向量训练

相较于 Twitter 等国外社交媒体平台,微博评论与其主要区别在于其用户使用的语言为中文,这导致微博评论相较于情感分类领域中英文文本的研究存在较大的差异性。首先,中文文本存在着分词问题,英文文本中的空格本身就是毫无争议的天然分词符,并且中文的分词是一个相对复杂的问题,往往依赖于语义信息及上下文信息;其次,微博评论中充斥着大量的流行语、外来语、谐音字、错用字以及网络用语,这些集外词(OOV)导致传统的情感分词算法很难有效地进行分词,而分词效果的优劣直接影响情感词向量的好坏,从而影响情感分类模型的性能。随着对深度神经网络,特别是卷积神经网络研究的不断深入,学者发现,即便将句子拆分为细粒度的单元,也能够依靠卷积神经网络的特征提取功能,学习出有效的语义。在英文文本领域中,已经有研究证明,即便完全抛弃英文单词,将英文分割成字母粒度,也能取得不错的文本分析效果[12]。此种方法,将英文拆分成理论上的最小单元,依靠卷积层的不断地特征提取及组合,最终形成有效的语义。基于以上特点,本文引入以“字”为单位的向量表示,从而最大程度保留微博评论的原始语义,将集外词分解为字,依靠卷积层,对字级别向量与词级别向

量进行有效的特征提取，防止因分词导致的信息丢失对下游模型的干扰。

随着深度学习在舆情情感分析领域的广泛应用，越来越多的研究者开始构建基于概率的浅层神经网络模型。这种模型训练出来的词向量，与语言模型高度相关。通过神经网络的逐层计算，模型最终输出一个低维度的向量，而词汇所蕴含的语义被分散地存储于向量的各个维度之中。这种构建方式，降低了词汇本身的维度，避免了独热编码等模型的维度灾难问题。在这些浅层神经网络模型中，谷歌公司提出的分布式词向量模型 Word2Vec[18] 是目前主流的词向量表示算法。根据对训练语料库中所构建词汇的上下文语义不断调节词向量，使得语料库中具有相似上下文的、语义相近的词汇具有相近的词向量表示[19]，本文选用 Word2Vec 模型对词向量进行建模。

### 1.3 卷积神经网络

卷积神经网络（Convolutional Neural Network，CNN）通常运用在图像处理领域，常用作图像特征的提取。随着研究的不断深入，卷积神经网络也逐渐被应用于文本特征的选取[20]。文本数据对应着海量权重，通过卷积神经网络的卷积层与池化层的运算，可以有效地进行特征选择[21]。从情感倾向性分析的角度来讲，这可以剔除文本中的冗余信息，将关键情感提取出来。

卷积神经网络中主要包含卷积和池化两个过程[22]：卷积的作用是为了特征选择，将自学习的“卷积核”以滑动窗口的形式扫过整篇文档（此处为表示方便，将情感文本中的文字转换为数字），整个运算过程即为卷积过程。卷积过程中重要的语义信息在运算过程中会被放大，非重要的语义信息将会被缩小；池化过程则是一个特征降维的过程，目的是减少计算量。在情感倾向性分析中，通常选择最大池化以保留最关键的语义信息。有了卷积和池化，卷积神经网络可以在提取出情感文本中重要特征的同时，大幅度降低维度，从而增加计算便利性。

## 2 基于字词向量的多尺度卷积神经网络微博评论情感分类模型

### 2.1 情感分类模型的构建过程

基于上述相关理论基础，本文提出基于字词向量的多尺度卷积神经网络情感分类模型（Multiscale Convolutional Neural Network，Multi-CNN），见图 1。模型构建顺序依次为微博评论数据采集及预处理、字词级别向量训练及选择、卷积神经网络训练、性能评测及模型选择、微博评论情感分类和数据分类结果分析等六个步骤：①微博评论数据采集及预处理，主要是通过网络爬虫的方式获取微博评论数据，并利用 Jieba 分词工具以字和词为最小分词单位进行分词，并进行去停用词操作的预处理；②字词级别向量训练及选择，运用 Word2Vec 训练词向量，并通过余弦相似度选取最优词向量，最后结合字向量，构建三维文本矩阵；③卷积神经网络训练，主要通过卷积层的特征提取，池化层的降维，将三维文本矩阵映射为一维向量，以供全连接层进行操作，然后通过模型训练，依靠反向传播算法在训练集上不断调节模型参数权重，并通过交叉验证进行情感词选择；④性能测评及模型选择，通过计算正确率、召回率、F1 值进行对比验证，并选用综合表现最优的模型作为最终分类模型；⑤微博评论情感分类，模型依据最后确定的最优模型进行微博评论的情感分类；⑥分类结果分析，最后对确定的微博情感分类结果进行讨论和分析。

### 2.2 文本向量训练与选择

在经过数据采集及预处理之后，微博情感分类需要解决的主要问题是将情感文本进行向量化表示，使之变为可计算的数据，以供情感分类模型使用。一般的解决方式是使用词嵌入模型，将每一个词映射为一个低维向量，从而使得情感文本最终被表示成一个由这些向量构成的矩阵[23]。本文在研究过程中使用谷歌开源的 Word2Vec 词向量训练工具，将每个字或经过分词处理后的词映射为一个向量，并使用余弦相似度作为评价词级别向量的指标，如公式(1)所示：

$$\mathrm{sim}(x,y) = \cos\theta = \frac{\vec{x}\cdot\vec{y}}{\|x\|\ \|y\|} \qquad \text{公式(1)}$$

对于词级别向量的优劣判断的思路是，在优质的词向量中相似的词汇所映射成的向量余弦值较高；反之无关的词汇余弦值较小。本文依据余弦相似度选取两组质量最优的词级别向量，再结合一组字级别向量构建三维文本矩阵，见图 2。

### 2.3 整体架构与算法流程

情感文本构建成三维文本矩阵后，通过卷积神经网络进行文本特征提取与降维。如图 3 所示，文本矩阵的起始维度为 200，经过卷积层处理后，虽能有效的提取文本特征，但矩阵的维度仍为 198，并没有明显地减少。经过池化层后矩阵的维度迅速降为 64，这正符合卷积神经网络不同操作的功能，即卷积操作目的是提取文本特征，池化操作目的是控制维度。卷积层与池化层交替使用，使得模型能够在有效提取文本特征的同时降低维度。

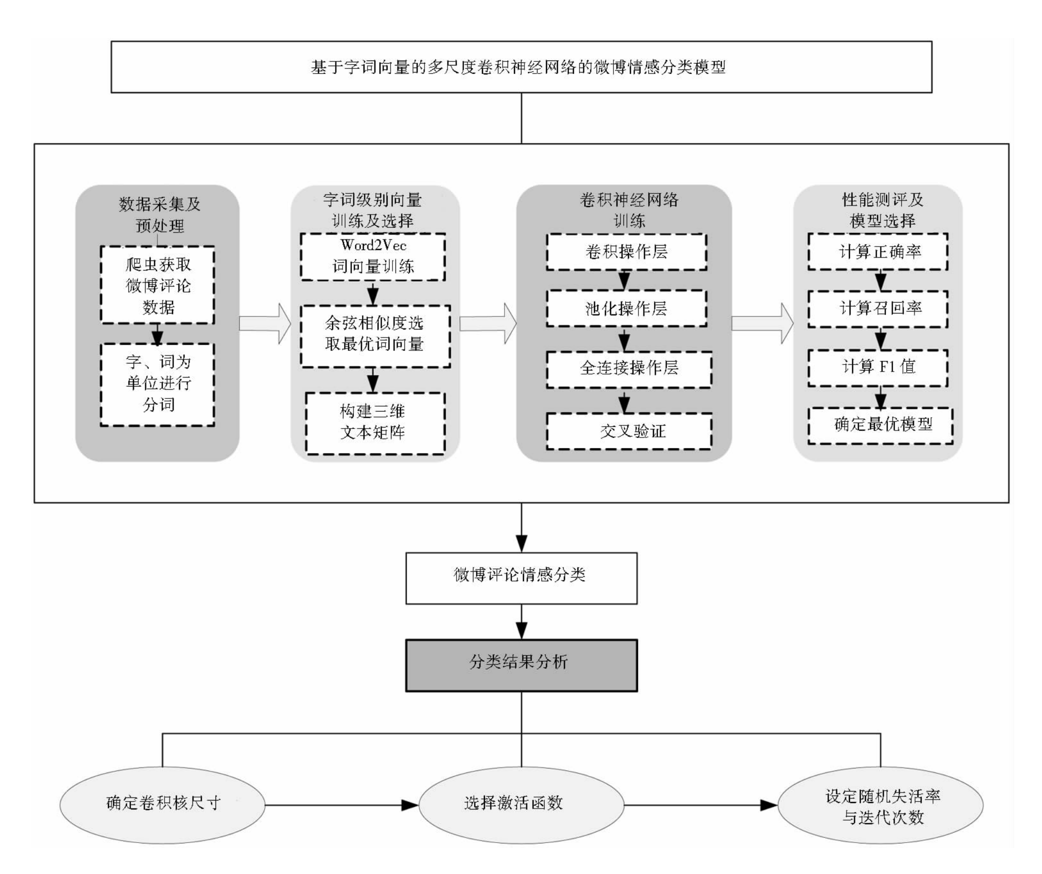



图 1　基于字词向量的多尺度卷积神经网络的微博评论情感分类模型

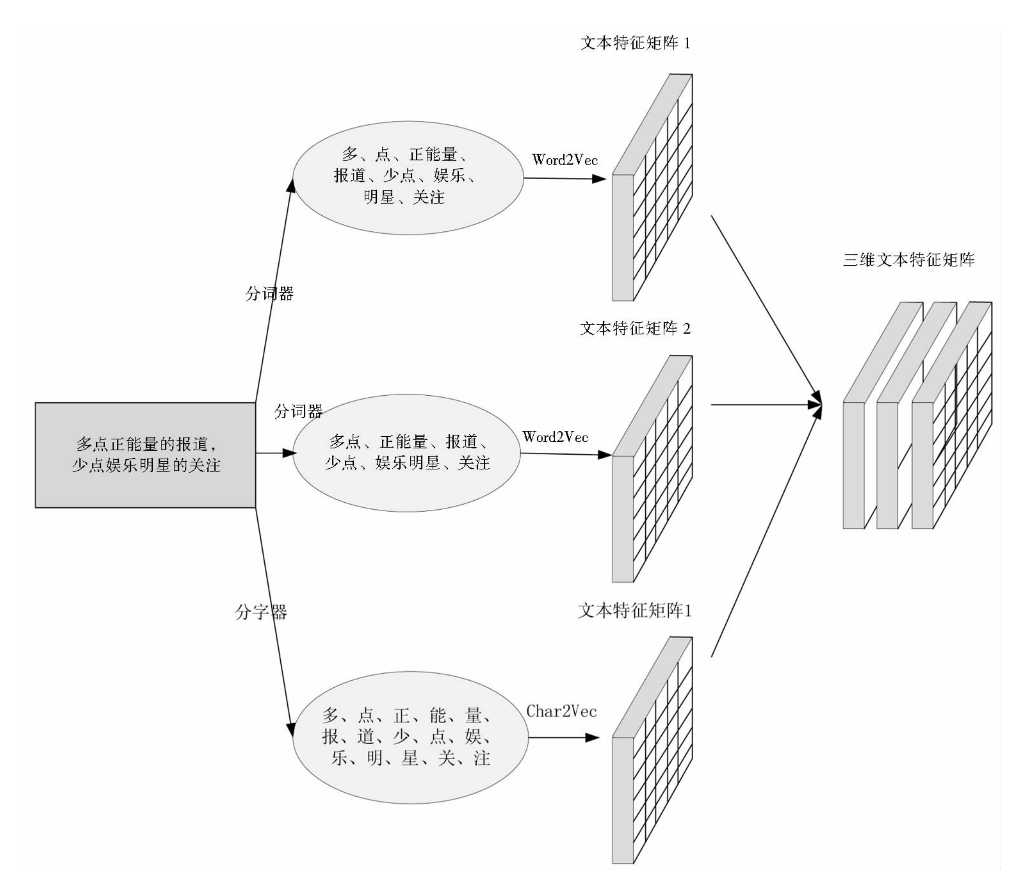



图 2　三维文本矩阵的构建示例

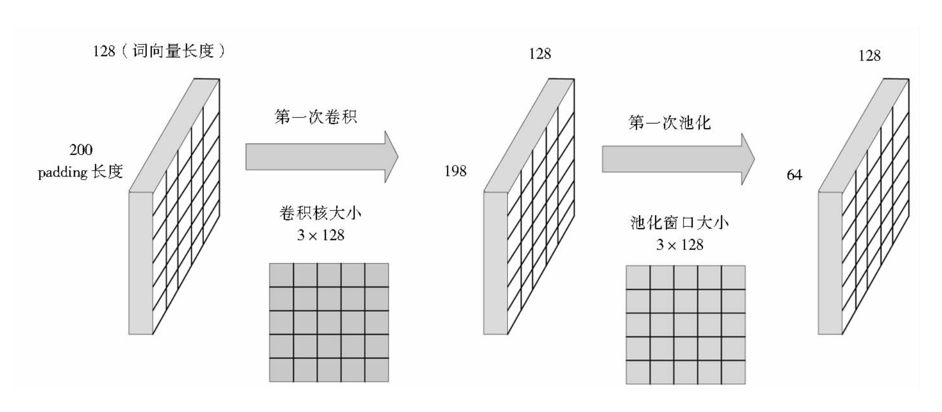



图 3　卷积操作与池化操作过程示意

如图 4 所示，本文依据 AlexNet 分组卷积设计了三种不同尺度的卷积核，分别为 3×128、4×128 和 5×128，以便于对情感文本中相邻的三、四或五个词进行文本特征提取。每种卷积核数量设置为 128 个，结合相同的最大池化操作形成卷积特征提取层。最终，将三种卷积单元结合成一个一维向量。

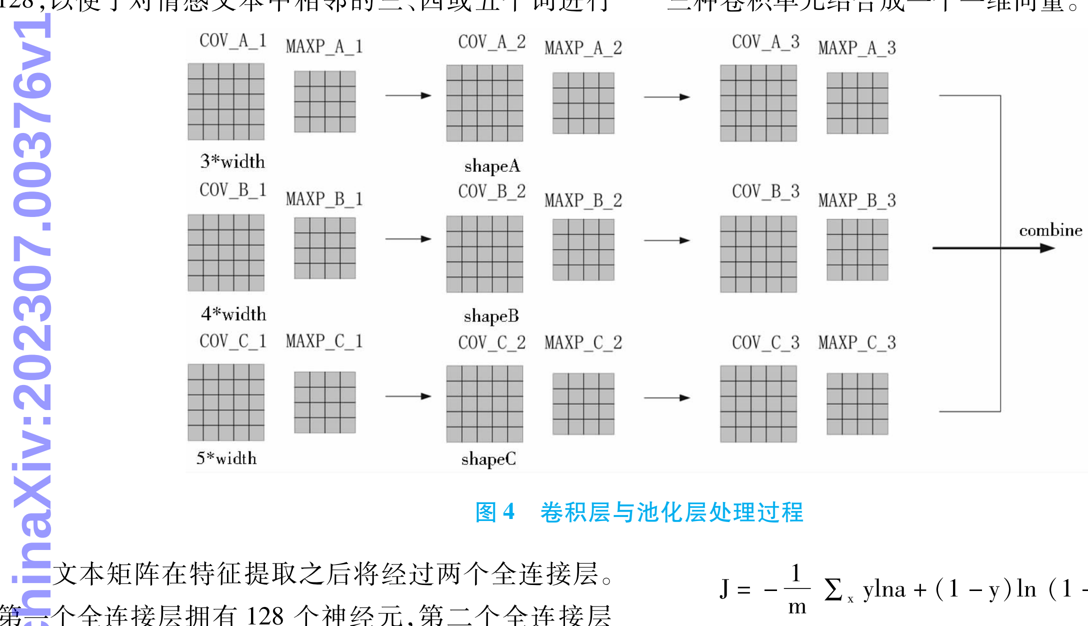



图 4　卷积层与池化层处理过程



文本矩阵在特征提取之后将经过两个全连接层。第一个全连接层拥有 128 个神经元，第二个全连接层拥有 64 个神经元，这样可以将特征提取后的文本矩阵重新拉抻为一个二维向量。因为全连接层的使用会大大增加模型的参数量，甚至造成过拟合，所以要加入 Dropout 层以控制全连接层神经元的随机失活数量。Dropout 在每轮迭代中会随机的将一定比例的神经元“杀死”，即指定其输出为零，使得其连接的权重不会在反向传播的训练过程中参与权重更新。这样一方面控制了模型的参数量，方便了计算；另一方面也割裂了某些权重之间的非必要依赖，降低模型过拟合的风险，使得模型在测试集上的准确率显著提升。

卷积神经网络训练阶段的本质是通过计算损失函数，并按照反向传播算法不断进行权重更新的迭代过程[24]。本文选择二分类交叉熵损失函数作为代价函数，损失函数的表达式为公式(2)：

$$J = -\frac{1}{m}\sum_x y\ln a + (1-y)\ln(1-a) \qquad \text{公式(2)}$$

其中，y 为期望输出，a 为神经元的实际输出，整体损失为 J。交叉熵损失函数具有良好的函数性质：①当期望与真实输出接近时(即期望与真实输出同为 1 或同为 0)，损失 J 接近为 0。②当期望与真实输出相距较远(即期望为 0 真实输出为 1，或期望为 1 真实输出为 0)，损失接近无穷大。由于神经网络的参数更新依赖于梯度下降，较大的损失意味着模型参数的更新，而损失为零则表明模型收敛，即期望与真实输出一致。

梯度计算为公式(3)：

$$\frac{\partial J}{\partial w} = \frac{1}{n}\sum_x x(\sigma(z)-y) \qquad \text{公式(3)}$$

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad \text{公式(4)}$$

其中，w 表示权重，z 表示每一层神经元的输入。其中，权重更新没有 $\sigma'(z)$ 这一项，仅依赖 $\sigma(z)-y$，误差较大时权重更新快，误差较小时权重更新慢。

在确定损失函数后，模型需要选择优化算法迭代求解最优值，本文选择目前深度学习领域广泛使用的 Adam 优化算法进行迭代更新[25]。最后，模型整体流程如图 5 所示。该过程主要分为三个部分，分别为文本向量化层（embedding 层）、卷积特征提取层和分类层。

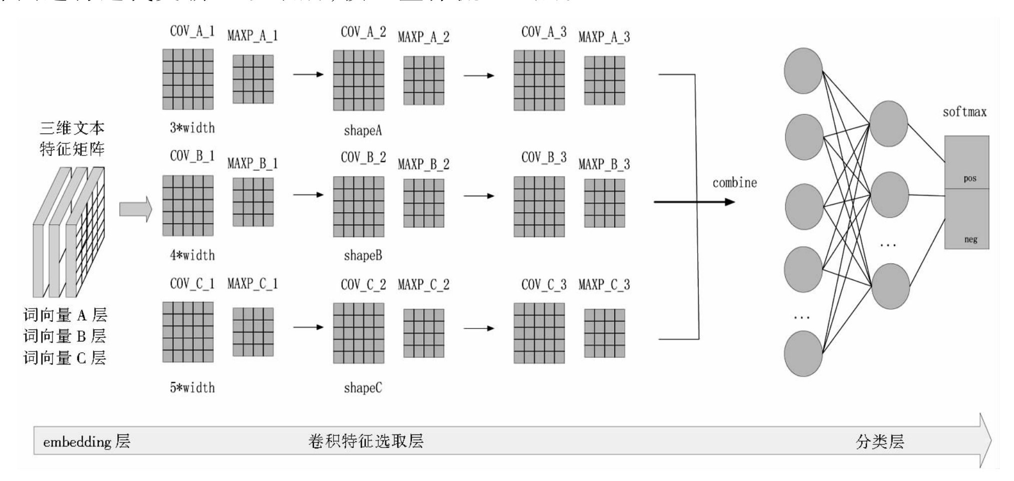



图 5　进行深度学习迭代算法流程

## 3　实验验证

### 3.1　数据预处理

为验证本文构建的微博情感分类模型的准确性，本文进行了对比性实验验证。本文以新浪微博中的“微博热搜整改”话题为例，使用 Python3.6，采用网络爬虫方式采集获取微博用户评论数据。获取的数据字段包括用户 ID、昵称、个人资料、微博内容、移动工具端类型、转发评论、点赞数量和评论时间等。根据百度指数的统计数据，关键词“微博热搜整改”舆情的生命周期为 2018 年 1 月 26 日至 2018 年 2 月 28 日，如图 6 所示。数据到达顶峰的时间段位于 1 月 27 日，因此本文选择该网络舆情的第一个高峰作为情感倾向性分析的研究起始点，以 2 月 28 日作为终止点。在“微博管理员”用户下采集微博用户评论信息。

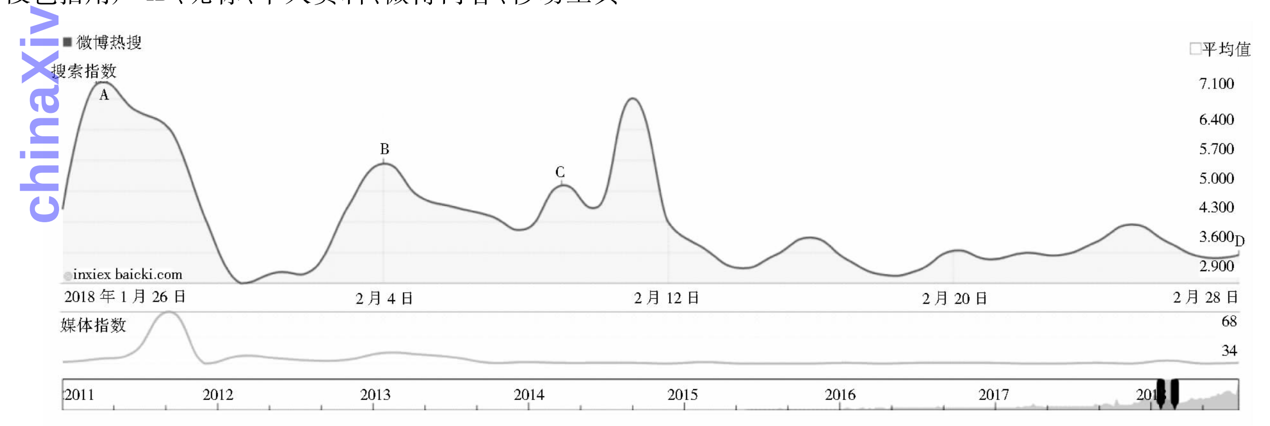



图 6　“微博热搜”百度指数（资料来源：http://index.baidu.com/）

原始微博评论数据包含商业广告、评测活动等噪声数据，并非所需要的情感文本数据。为消除这部分噪声数据对实验效果的影响，在中文分词前，需要进行数据清洗。本文选用的数据清洗方法为基于关键词的识别方法，并通过分析总结出这部分噪声数据大多数包含相同的主要关键词，对包含该类关键词的评论数据进行剔除，去除评论数据中仅“@ 其他微博用户”的评论以及“http”开头的锚链接，形成初步的数据集，共计 134 765 条，经初步筛选后，剩余 116 764 条，并招募在校大学生对其进行正负情感分类，共计招募大学生 175 人，采用投票法对初步筛选后的 116 764 条数据进行人工二分类，统计分类结果后，票数差大于 25 票的视为有效数据，原因是为了保证正负样本的明显倾向性。例如：若某一样本为正样本，其得到的正样本的投票数至少为 100 票，负样本的投票数则至多为 75 票。设定票差是为了从总体样本中筛除具有歧义性的样

本，保证样本的明显倾向性，从而有助于模型的训练。经过筛选后最终确定 91 376 条数据。其中正向情感 38 185 条，负向情感 53 191 条。

### 3.2 实验内容设计

本文设计了三个实验：

首先，为验证引入字级别向量对模型效果的提升，选取相同的语料数据训练维度相同的字向量与词向量各两组，以验证引入字级别向量对模型的帮助。

其次，为了确定模型超参数，并进一步验证基于字词向量的多尺度卷积神经网络微博评论情感分类模型的有效性及优越性，本文对实验内容及流程进行了如下设计：①训练/测试集划分。每次实验随机选取 80% 的数据集作为训练集，选取剩余的 20% 作为测试集，测试集不参与训练，仅用来验证模型性能；②文本向量训练及选择。为了保证文本向量的质量，本文使用搜狐全网新闻数据语料库训练词向量。搜狐全网新闻数据语料库搜集了国内若干新闻站点 2012 年 6 月 -7 月期间的 20 000 余篇真实新闻数据，涵盖社会、体育、新闻、娱乐等 18 个频道，数据真实，覆盖领域广。选择该语料库进行词向量训练，可以尽量降低词向量的针对性，增强泛化能力；③构建卷积神经网络模型。采用 Keras 深度学习工具包和 TensorFlow 作为其后端支持。④超参数调节。为了提升模型分类性能，对卷积核尺寸、激活函数、dropout 随机失活比率以及迭代次数进行调节；⑤效果对比。

最后，本文选用国家语言资源监测与研究中心发布的"2017 年度十大网络用语"[26] 作为集外词。统计数据集中出现这些集外词的相关评论信息，选取数目最多的网络用语，构建数据集，计算不同模型对其的分类正确率，验证模型对集外词的处理能力。



## 4 讨论分析

### 4.1 中文字向量分类能力实验

实验使用相同的语料数据训练维度相同的字向量与词向量各两组，均为 128 维。选取数据集中的 10 000条数据，其中正负样本各计 5 000 条，构建数据子集。同时，搭建两种形状结构相同的卷积神经网络，卷积核形状均为 3×128，设置三个该形状的卷积核，结合同样形状的池化层，最后一层为单神经元的全连接层。并选取逻辑斯蒂回归及高斯核 SVM 分类器作为另外两个模型，训练并观察其分类结果。结果如表 1 所示：

表 1 中文字向量分类能力

| 模型 | 词向量 1 | 词向量 2 | 字向量 1 | 字向量 2 |
|---|---|---|---|---|
| CNN 准确率 | 0.761 1 | 0.780 2 | 0.809 4 | 0.801 0 |
| SVM 准确率 | 0.731 3 | 0.741 7 | 0.723 1 | 0.703 2 |
| Logistic 准确率 | 0.723 4 | 0.750 1 | 0.761 9 | 0.724 2 |

实验结果显示，中文字向量在结合卷积神经网络的特征提取后，能够带来一定分类能力的提升。同时，在其他模型上，中文字向量也有着不错的分类效果。

### 4.2 基于字词向量的多尺度卷积神经网络超参数确定及对比实验

在经过词向量训练后，需要通过超参数的调节以确定最优分类模型。超参数不同于模型本身的权重参数，无法通过梯度下降的方式进行调优。这些超参数的不同选择，在很大程度上影响了模型整体的性能。本文针对卷积核尺寸、激活函数、Dropout 随机失活率及迭代次数为待调整的超参数，通过实验进行讨论分析，以确定最优的超参数组合。

4.2.1 *卷积核尺寸* 卷积神经网络中卷积核尺寸是一个重要参数。由于情感分类问题中使用的是一维卷积核，只需验证卷积核的长度对模型性能的影响，卷积核的长度意味着滑动窗口将关注多少个近邻词汇[27]。本文考虑相邻的 2-6 个词汇，共五种卷积核，结合三层卷积的网络架构，设计了六种结合方式并进行实验对比，如表 2 所示。

表 2 选用不同卷积核尺寸的实验对比表

| 各层卷积核尺寸 | 训练集正确率 | 测试集正确率 | 召回率 | F1 值 |
|---|---|---|---|---|
| 2,3,4 | 0.981 7 | 0.872 0 | 0.943 6 | 0.913 0 |
| 4,3,2 | 0.977 0 | 0.875 6 | 0.942 2 | 0.912 9 |
| 3,4,5 | 0.971 2 | 0.877 6 | 0.941 8 | 0.912 8 |
| 5,4,3 | 0.964 3 | 0.862 1 | 0.951 4 | 0.910 1 |
| 4,5,6 | 0.991 4 | 0.872 2 | 0.945 9 | 0.911 9 |
| 6,5,4 | 0.951 7 | 0.843 0 | 0.916 2 | 0.893 1 |

实验结果表明，各层卷积核尺寸选为 3,4,5 时，模型在测试集的分类正确率最高，能够得到分类精度较高的模型；选为 2,3,4 时，模型的召回率与 F1 值最高，能够得到正例识别率较高的模型；与此同时，卷积核尺寸大于或等于 5 时，模型产生了过拟合，在测试集上精度会有明显下降。实验结果符合人们关于语言的主观认识，即在一般情况下，词之间的距离超过五时，其相互作用往往几乎不存在，也就无法利用这种尺寸的卷积核提取有效的语义特征。

4.2.2 *激活函数* 神经网络的激活函数是其非线性变换的重要组成部分。当今深度神经网络中常见的激

活函数一般有 tanh 与 ReLU 两种[28]。计算方式如公式(5)、公式(6)所示：

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad \text{公式(5)}$$

$$\mathrm{ReLU}(x) = \max(0, x) \qquad \text{公式(6)}$$

本文使用测试集上表现最优的卷积核尺寸进行调节卷积层与全连接层的激活函数。如表 3 所示。实验结果显示，在卷积层使用 tanh 函数，全连接层使用 ReLU 函数时，模型综合表现最优。其中，若卷积层选用 ReLU，本身会破坏神经网络中初始化的权重，若是在此种情况下将全连接层的激活函数选为 tanh，则经过若干轮迭代后，在前向传播过程中双曲正切函数的输入将都会集中在 x 轴正半轴，双曲正切函数在无穷远处的梯度值为零，将导致模型出现严重的梯度弥散问题。

表 3　选用不同激活函数在测试集的实验对比结果

| 测试 | 卷积层 | 全连接层 | 测试集正确率 | 测试集召回率 | 测试集 F1 值 |
|---|---|---|---|---|---|
| 测试 1 | tanh | tanh | 0.871 5 | 0.932 4 | 0.912 4 |
| 测试 2 | tanh | ReLU | 0.894 2 | 0.923 3 | 0.927 2 |
| 测试 3 | ReLU | ReLU | 0.867 8 | 0.898 2 | 0.903 4 |

4.2.3　Dropout 随机失活率与迭代次数　网络结构中 Dropout 层的随机失活率以及训练的迭代次数也是影响模型性能的重要因素。以下为相关实验的验证过程。首先测试 Dropout 随机失活率对模型性能的影响，在测试集上的相关指标如下图 7 所示，可以看出 Dropout 随机失活率对模型的影响趋势。当 Dropout 随机失活率处在 0.05 –0.1 时，模型的各项指标达到最优效果；当随机失活率超过 0.1 后模型的整体指标呈现下降的趋势。因此，将 Dropout 随机失活率设置为 0.05 –0.1 之间可获得最好的模型性能。

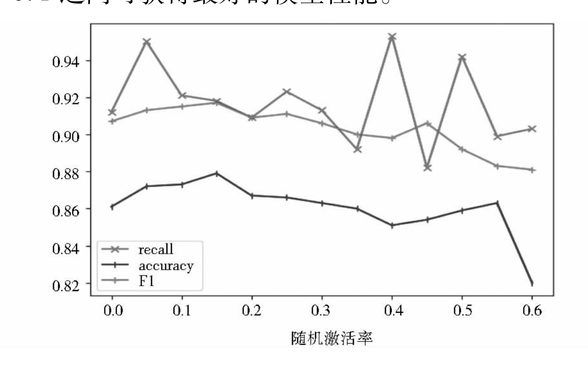



图 7　Dropout 随机失活率对模型性能的影响

深度神经网络中的训练迭代次数需要人工调整超参数，迭代次数的设定往往因所处理任务的不同而不同。迭代次数过少，会使模型不能够收敛到局部极小值而导致欠拟合现象；反之迭代次数过多，一方面延长了模型的训练时间，另一方面也使得模型面临过拟合问题，失去应有的泛化能力。图 8 显示了迭代次数分别为 1 –20 时，模型相关指标的变化情况。从图 8 中可以看出，在迭代 15 轮以后，模型的各项指标已经达到最优效果；16 轮之后各项指标有明显下滑趋势，表明此时模型已经过拟合。因此，本文将迭代次数设定在 15 次。

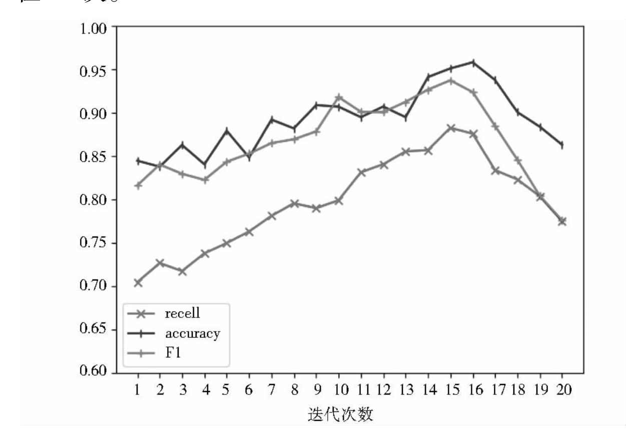



图 8　迭代次数对模型相关指标的影响

4.2.4　实验结果对比　为验证本文所构建模型的优越性，本文选择了目前主流的情感分类模型在相同数据集上进行了实验结果对比。实验对比结果如表 4 所示：

表 4　情感分类模型对比实验结果

| 模型 | 正确率 | 召回率 | F1 值 |
|---|---|---|---|
| CBOW-SVM | 0.784 4 | 0.772 4 | 0.773 8 |
| Word2Vec-CNN | 0.851 6 | 0.8499 | 0.850 1 |
| Att-CNN | 0.902 9 | 0.891 8 | 0.874 9 |
| Multi-CNN | 0.947 3 | 0.923 7 | 0.927 4 |

其中，CBOW-SVM 是直接将文本数据通过 N-gram 模型输入到 SVM 中进行分类；Word2Vec -CNN 是利用 Word2Vec 训练的词向量，利用卷积神经网络进行训练；Att-CNN 则是利用 LSTM 为基础架构的引入注意力机制的模型对词向量组成的文本矩阵进行训练。对比实验数据结果，表明本文提出的 Multi-CNN 模型在微博评论这种短文本分类任务中的各项指标都相对较高，并表现出较准确的分类效果。

### 4.3　基于字词向量的多尺度卷积神经网络对集外词的处理能力

为验证模型对集外词的处理能力，本文选用国家语言资源监测与研究中心发布的“2017 年度十大网络用语”作为集外词。2017 年度十大网络用语如表 5 所示：

表 5　2017 年度十大网络用语

| 序号 | 网络用语 |
|---|---|
| 1 | 打 call |
| 2 | 尬聊 |
| 3 | 你的良心不会痛么 |
| 4 | 惊不惊喜，意不意外 |
| 5 | 皮皮虾，我们走 |
| 6 | 扎心了，老铁 |
| 7 | 还有这种操作 |
| 8 | 怼 |
| 9 | 你有 freestyle 么 |
| 10 | 油腻 |

统计数据集中出现这些集外词的相关评论信息，其中"还有这种操作"数目最多，达到 1 146 条，其中正向情感数目为 487 条，负向情感数目为 659 条。统计 4.2.4 模型对其分类的正确率，结果如表 6 所示：

表 6　情感分类模型正确率对比

| 模型 | 正确率 |
|---|---|
| CBOW-SVM | 0.534 |
| Word2Vec-CNN | 0.751 |
| Att-CNN | 0.769 |
| Multi-CNN | 0.873 |

可见，由于集外词的存在，CBOW 这类基于 N-gram 的模型很难构建出有效的语义特征，其分类效果较差。CNN 类模型由于卷积层的存在，能够有效的进行特征提取，其分类效果远远高于 CBOW。而 Multi-CNN 由于采用了更为细粒度的字级别向量，其正确率明显超过了其它 CNN 模型。从而该实验数据证明了其对集外词具有良好的处理能力。

## 5　研究结论

在理论层面，本文基于现有微博评论由于上下文信息有限、集外词多等情感分类的难题，提出了基于字词向量的多尺度卷积神经网络微博评论的情感分类模型。该模型运用卷积层来学习微博评论中的结构特征及语义特征；同时运用多尺度卷积核来捕捉词语间的更多联系，以有效关注不同程度的上下文信息，从而改善微博评论中上下文信息有限的条件制约；并选用字词向量相结合的方式，在一定程度上减轻了集外词对模型情感分类性能的影响。通过爬取微博评论的真实数据，验证了模型的可行性，该模型可为微博评论情感分类模型的构建及方法研究提供新的理论模型和方法支撑。

在实践层面，本文通过实验方法对影响模型性能的超参数提供了设置思路，并对比了主流的情感分类的相关研究方法，构建和验证微博评论情感分类模型。通过运用卷积神经网络，可以分析网络用户对话题事件的关注程度和情感变化，从而为相关舆情监管部门正确管控话题事件的发展状况和舆情走向起到一定的指导作用，为微博舆情的动态调控及监测提供更为可行性的管理办法。

本文选用字词向量结合多尺度卷积核的方式在一定程度上解决了集外词等问题，但在研究过程中仍然存在一定的缺陷，卷积神经网络训练时间较长，并且对超参数的设置十分敏感。在后续的研究中，将考虑通过一些正则化方式降低模型的过拟合风险及模型的训练难度，从而进一步提升模型的泛化能力。

**参考文献：**

[1] 新浪科技. 微博发布 2018 年第二季度未经审计财报[EB/OL]. [2018－10－08]. https://tech.sina.com.cn/i/2018-08-08/doc-ihhkuskt9159883.shtml.

[2] NEVES-SILVA R, GAMITO M, PINA P, et al. Modelling influence and reach in sentiment analysis[J]. Procedia CIRP, 2016, 47:48－53.

[3] SOLEYMANI M, GARCIA D, JOU B, et al. A Survey of Multimodal Sentiment Analysis[J]. Image & vision computing, 2017,65(9):3－14.

[4] ÖZTURK N, AYVAZ S. Sentiment analysis on Twitter: a text mining approach to the Syrian refugee crisis[J]. Telematics & informatics, 2018,35(4):136－147.

[5] 唐晓波，兰玉婷. 基于特征本体的微博产品评论情感分析[J]. 图书情报工作，2016(16):121－127.

[6] 马松岳，许鑫. 基于评论情感分析的用户在线评价研究——以豆瓣网电影为例[J]. 图书情报工作，2016(10):95－102.

[7] 梁晓敏，徐健. 舆情事件中评论对象的情感分析及其关系网络研究[J]. 情报科学，2018,V36(2):37－42.

[8] 唐慧丰，谭松波，程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报，2007,21(6):88－94.

[9] 张冬雯，杨鹏飞，许云峰. 基于 Word2vec 和 SVMperf 的中文评论情感分类研究[J]. 计算机科学，2016,43(6):418－421.

[10] NIHAR M. R, RAJESH S. P. LFNN: Lion Fuzzy Neural Network-based evolutionary model for text classification using context and sense based features[J]. Applied soft computing, 2018,71:994－1008.

[11] 冯兴杰，张志伟，史金钏. 基于卷积神经网络和注意力模型的文本情感分析[J]. 计算机应用研究，2018,35(5):1434－1436.

[12] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification [J]. Computer science, 2015(9):1－9.

[13] 庞俊. 基于确定话题和情感极性的博客文本聚类研究[D]. 武汉：武汉理工大学，2010.

[14] 王晰巍,张柳,文晴. 基于贝叶斯模型的移动环境下网络舆情用户情感演化研究——以新浪微博"里约奥运会中国女排夺冠"话题为例[J]. 情报学报,2018,37(12):1241-1248.

[15] ZHANG Q, GAO Z, LIU J. Research of Weibo short text classification based on Word2vec[J]. Netinfo security, 2017(1):57-62.

[16] 周瑛,刘越,蔡俊. 基于注意力机制的微博情感分析[J]. 情报理论与实践,2018,41(3):89-94.

[17] 控杏,林庆. 主观性文本情感分类研究综述[J]. 信息技术,2018(8):126-130.

[18] Word2vec. [EB/OL]. [2018-10-08]. https://en.wikipedia.org/wiki/Word2vec.

[19] 周练. Word2vec 的工作原理及应用探究[J]. 图书情报导刊, 2015(2):145-148.

[20] 卢宏涛,张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理,2016,31(1):1-17.

[21] 吴鹏,刘恒旺,沈思. 基于深度学习和 OCC 情感规则的网络舆情情感识别研究[J]. 情报学报,2017,36(9):972-980.

[22] REN Y, JI D, REN H. Context-augmented convolutional neural networks for twitter sarcasm detection[J]. Neuro computing, 2018,308(9):1-7.

[23] 于政. 基于深度学习的文本向量化研究与应用[D]. 上海:华东师范大学,2016.

[24] 王毅,谢娟,成颖. 结合 LSTM 和 CNN 混合架构的深度神经网络语言模型[J]. 情报学报,2018(2):194-205.

[25] SMITH L N. Cyclical learning rates for training neural networks [J]. Computer science, 2017:464-472.

[26] 2017 年度十大网络用语. [EB/OL]. [2019-02-22]. https://baike.baidu.com/item/2017 年度十大网络用语/22273910? fr=aladdin.

[27] 陈珂,梁斌,柯文德,等. 基于多通道卷积神经网络的中文微博情感分析[J]. 计算机研究与发展,2018,55(5):945-957.

[28] 曲之琳,胡晓飞. 基于改进激活函数的卷积神经网络研究[J]. 计算机技术与发展,2017,27(12):77-80.

**作者贡献说明:**

张柳:负责论文撰写、修改及数据采集、数据处理;

王晰巍:负责研究命题及研究思路的制定,论文撰写及论文终稿修订;

黄博:负责数据处理及技术支撑;

刘宇桐:负责数据处理及摘要翻译。

## A Sentiment Classification Model and Experimental Study of Microblog Commentary Based on Multivariate Convolutional Neural Networks Based on Word Vector

Zhang Liu[1] Wang Xiwei[1,2] Huang Bo[3] Liu Yutong[1]

[1] School of Management, Jilin University, Changchun 130022

[2] Research Center for Big Data Management, Jilin University, Changchun 130022

[3] School of Computer Science and Technology, Jilin University, Changchun 130022

**Abstract:** [**Purpose/significance**] The microblog commentary sentiment classification model can play a guiding role for the relevant public opinion supervision departments to correctly control the development of the topic events and the public opinion. [**Method/process**] Based on the multi-scale convolutional neural network of word vector, this paper used multi-scale convolution kernel to improve the conditional constraints of finite context information in microblog commentary, and constructed multi-scale convolutional neural network microblog commentary emotion classification model based on word vector. Finally, the feasibility and superiority of the model were verified by crawling the real data of "microblogging hot search and rectification". [**Result/conclusion**] Verification results show that the multi-scale convolutional neural network based on word vector performs well in the short text classification task with limited context information such as weibo public opinion. On the theoretical level, this paper provides a more accurate emotional classification theory model and classification method for microblogging public opinion emotion classification. In practice, it can better guide the public opinion supervision department to better guide and supervise the emotional sentiment of public opinion.

**Keywords:** convolution neural network word vector multiscale convolution kernel Weibo comments sentiment classification